The Report Committee for Yang Yang
Certifies that this is the approved version of the following professional report:

An Exploratory & Quantitative Analysis on the Correlations between Urban Form and Urban Vitality in Downtowns of Three Main Cities in the Texas Triangle

APPROVED BY
SUPERVISING COMMITTEE:

Junfeng Jiao, Supervisor

Robert Paterson

An Exploratory & Quantitative Analysis on the Correlations between Urban Form and Urban Vitality in Downtowns of Three Main Cities in the Texas Triangle

by

Yang Yang

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Community and Regional Planning

The University of Texas at Austin

August 2019

Acknowledgements

I would like to acknowledge my supervisors Professor Junfeng Jiao and Robert Paterson for providing me their insights and knowledge about urban form and vitality, and critiques in my professional report and supporting me to finish the report. Without their support, the report would not have been possible.

I would also like to thank my friends, especially Bolin Lan and Liang Chen for reading my report and sharing their opinions with me. Finally, I would like to thank my classmate, Madison Graham to help me to proofread my report. All of their help would be grateful for the report.

Abstract

An Exploratory & Quantitative Analysis on the Correlations between Urban Form and Urban Vitality in Downtowns of Three Main Cities in the Texas Triangle

Yang Yang, M.S.C.R.P

The University of Texas at Austin, 2019

Supervisor: Junfeng Jiao

Most previous studies on vitality were conducted through qualitative method. Though there is a new trend that conducting quantitative research on relationship between urban form and urban vitality, with many new breakthroughs in recent technological advancements, there is still a lack of studies on urban vitality from a perspective of social media data. This report focuses on an exploratory analysis of how impacts of urban form effect on urban vitality through a quantitative method. It particularly explores influence of the six identified metrics of urban form on Yelp reviews as a proxy for urban vitality in 23 block groups within the three main cities of the Texas Triangle. The six metrics are also categorized into three types of elements, including density, urban layout, and street layout design. The analysis of the study combines a literature review and a multi-level and multivariate regression model. The aim of the report is to explore the influence of each urban form indicator on urban vitality and propose policies and guideline for future development.

The result indicates that out of the three elements of urban form, density has the most significant association with urban vitality. Activity density, connected node ratio and average block length are strongly correlated with vitality. The report can be served as a start point for a more comprehensive future research.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Background

Urban Vitality has been deeply discussed by urban planners and designers. Many of them have been attempting to define vitality and are deliberating on factors that affect it. Urban morphology, the form of human settlement, is one of the most important factors that influence vitality. However, previous studies showing vitality in relation to urban form are mostly conducted through qualitative research, such as field observations and questionnaires. Although some researchers have started exploring the relationship by quantitative methods, most researchers focus on a limited element of urban form. For instance, Zumelze studies the effects of urban form on neighborhood vitality by identifying land use mix, block size, plot sizes and building adaptability as urban form indicators (Zumelzu, 2019). In previous studies, street connectivity of urban form has been always ignored. There are also few researchers studying the correlation between urban form and vitality through the application of location-based check-in data sourced from social media platforms, even though it has become an important part of our lives.

Since the Texas Triangle is one of the fastest-growing megaregions in the United States, this professional report will focus on the impact of urban forms (activity density in relation to land use mix, sidewalk density index, intersection density, connected node ratio, average block length and average block size) of three main cities' downtown (the City of Austin, the City of Dallas and the City of San Antonio) in Texas Triangle on urban vitality measured by total number of Yelp reviews. These urban form metrics will be categorized into three types of elements, including design, urban layout and street layout. The objective of the report is to explore the influence of each urban form indicator on urban vitality by developing a framework to quantify urban form and urban vitality indicators. Multi-level and multivariate regression models will be conducted, and will suggest a comprehensive guideline for urban planners and designers to future downtown development for the polycentric Texas Triangle based on the selected urban form factors.

The report is separated into seven chapters. The structure of the report is shown below. Following the introduction, Chapter 2 reviews the relevant literature on the concept and representation of urban vitality, five elements of urban form and their relation to vitality through existing qualitative and quantitative methods. Chapter 3 describes how the study will be conducted and what techniques will be employed in the study. The research framework is summarized with a diagram in this chapter. Chapter 4 provides a comprehensive analytical process, starting with study area identification, then defines and calculates metrics of urban form and urban vitality individually, and will conclude with processing regression models to study the relationship between them. The preliminary results on metrics of urban form and vitality along with detailed information on data sources are also included in this chapter. Chapter 5 summarizes the results of quantitative analysis about how three elements of urban form, including design, urban layout and street layout influence urban vitality, specifically how urban vitality is affected by each urban form variable. Chapter 6 discusses limitations of the research and suggests future research. Chapter 7 summarizes the conclusion and policy implications for future development in the Texas Triangle.

Research Questions:

1. What factors of urban form of the built environment in downtown areas of three selected cities in the Texas Triangle correlate with urban vitality?
2. What elements of urban form in downtowns of the three main cities in the Texas Triangle correlate with urban vitality?
3. How strongly are urban form variables correlating with urban vitality?

Chapter 2: Literature Review

Concept of Urban Vitality

The term, vitality, was first introduced by Jane Jacobs in her *the Death and Life of Great Cities*. In her view, urban vitality reflected all kinds of diversity which is generated by close interaction and social networks among people. In other words, Jacobs presented four primary conditions and two secondary conditions that create diversity from which good cities draw their urban vitality. The four primary conditions are mixed land use that attracts more than one type of people, small block sizes that promote walking activities, high density and diverse architecture with a mix of new and aged buildings, accessibility, and border vacuums (Jacobs, 1961). Later, Lynch defined vitality as one of the seven primary indicators to measure the quality of good city form. Based on the human-oriented dimension, an urban vitality depends on the degree to which the form of the settlement "supports the vital functions, the biological needs and abilities of human beings, and how it protects the survival of species" (Lynch, 1984). Unlike Jacobs, Lynch mainly focused on biological and ecological dimensions and assesses the vitality only through three principals: survival-availability of all elements to sustain life, safety and the environment's consonance with the basic biological structure of human being. (Lynch, 1984).

Similarly, Montgomery came up with the same ideas as Jacobs' on a diversity of transactions between people and enough mixture of uses that promotes a diverse range of activity happening in streets, buildings and spaces (Montgomery, 1995). Montgomery suggested that the diverse primary uses include tea houses, cafes, restaurants, delicatessens, bakeries, cinemas, galleries, pubs and clubs, which provide a means to fulfill needs of different levels of the urban population. Both Jacobs and Montgomery advocated that a high level of mixture of uses and a wide diversity of activity play an important role in urban vitality. In addition to the diversity and mixed-use that drive activity, Montgomery also suggested a full set of principles on form, including high development intensity and density, flexible building adaptability, walkable human scale, a short city block and incremental permeability, street contact, fine grain, public realm, public green and water space, and landmarks or visual stimulation as principals for a successful city form and urban vitality. There is no doubt that the public realm, public green and water space, and landmarks act as attractors for meeting places in public life. In order to make the urban population feel more comfortable to walk, Montgomery suggested that a system of a walkable scale short blocks should be the primary building type for development and adding more intersections, which will brings benefits to form a successful urban district, street life and then vitality (Montgomery, 1995).

Gehl also believed that urban vitality is derived from people and their social activities. His study mainly emphasized urban public spaces in relation to the desire of people. He concluded that the well-functioning pedestrian system, including scale and dimensions of streetscapes, mixed functions, active frontages and open block are important to vibrant public life (Gehl, 1971). Vitality has been mentioned many by New Urbanists. For instance, Katz explained pedestrian-oriented streets, compactness, walking scale, appropriate density and mixed-use are the main drivers of the urban vitality (Katz, 1993).

In summary, most urbanists believed that vitality is mainly generated by social interactions among various types of the urban population. In order to attract more social activities and vitality, it is necessary to note mixed land use, small block sizes, high density and diverse architecture (Jacobs, 1961), high development intensity, building adaptability, human scale, permeability, street contact, fine grain, public realm, pubic green space and water space, visual stimulation (Montgomery, 1995), well-functioning pedestrian system, active frontages and open block (Gehl, 1971) and survival, safety and consonance (Lynch, 1984).

Methodology and Measurement on Urban Vitality Analysis

Mainstream methods to assess vitality are either qualitative or quantitative methods. Some researchers have started to measure and analyze urban vitality by quantitative methods, but the data they collect is still conducted through qualitative methods, such as field observations or questionnaires. For instance, Zarin used a questionnaire survey to collect data from 384 men on factors influencing vitality in Golestan and Narmak streets in Tehran, Iran (Zarin, 2014). Compared to Wu, who assessed the level of vitality in the Shangdi-Qinghe subdistrict of Beijing, generated an activity diary through a more accurate questionnaire survey based on GPS tracking devices carried on 1113 respondents. The respondents were still asked to provide detailed information on activities due to unreliable data on indoor activities (Wu, 2017). Through these data, to a certain extent, an approximate level of vitality can be discovered, however, the data collection by field observations and questionnaires requires a considerable amount of effort and time (Lu, 2019). In addition, qualitative analyses may bias due to the relatively subjective nature of field observations and questionnaires. Meanwhile, some researchers such as Drewes and Aswegen argue that a city's vitality can be measured by its normative welfare indicators, such as GDP, quality of life (QOL) and other social indicators, such as population growth over a period of time (Drewes, 2010). Quality of life reflects a city's well-being, community life, political security, education and job security at a large scale of cities. While providing sufficient measurement for regional or national economies, it is difficult to distinguish these indicators from a small spatial scale.

In recent years, emerging new technologies such as Information and Communication Technologies (Kitchin, 2014), Geographic Information System (GIS), Global Positioning Systems (GPS) and open data on new sources, including mobile phone data and social media data offer a new trend of quantitative analysis on vitality. Urban researchers have attempted to use quantitative methods to identify some aspects of vitality. The quantitative analysis of vitality data sourced from mobile phone made it possible for Yue to determine the level of vitality in Shenzhen, China. She used the number of a mobile phone users in a 24-hour period as an indicator of urban vitality (Yue, 2016). In addition to mobile phone data Yue harnessed to quantify vitality, Zumelzu evaluated vitality by analyzing the intensity of spatial occupation and different usage patterns and human activities in space by using two observation techniques from the space syntax, including the Gate method and the Static Snapshot (Zumelzu, 2019). Sulis summarized the new dataset into three types of spatial data that include a smart card, mobile phone, and social media. Respectively, the smart card data was used for calculating commuting flows, the mobile phone data was used for calculating activities and identifying urban dynamics, and finally, the social media data was also used for calculating social activities and networks (Sulis, 2018). He quantitatively measured the vitality of 363 Chinese cities by calculating point of interest density (POID), degree of urban function mixing (MIX), location check-in density (CIQD), housing prices, and population change. The point of interest data comes from Baidu Map Open Interface Technology, the check-in data was collected from Sina Weibo, one of the most populated social media platforms in China, and the housing price information was derived from Anjuke, one of the largest web platforms on housing in China (He, 2018). He thoroughly took into consideration quantifying vitality by new technologies and open data. Her contribution to the field can be seen as a foundation or new solution for future research.

Recent works have started to quantitatively analyze urban vitality based on a wide range of data from different types of sources, of which point of interest data, MIX data and other user's data from mobile phones have been investigated heavily. With many new breakthroughs in recent technological advancements, there is still a lack of studies on urban vitality from a perspective of social media data. Nowadays, the social media platform has changed our lives drastically. Millions and millions of population snap, take selfies and share their thoughts on social media every day. Therefore, this research utilizes geo-located data from a social media platform to explore its representativeness of city vitality.

Urban Form and Its Relations to Urban Vitality

Urban form can be defined as the spatial pattern of fixed elements (Anderson, 1996). In general, it can be categorized into five interrelated elements that include density, urban layout, land use, housing and building characteristics and transport infrastructure. A number of physical features may be contained in these elements of urban form, including scale, density, size, shape, land uses, block layout, street connections, building types, and distribution of green space (Dempsey, 2009). A wide range of indicators can be used to represent each element. For instance, in order to determine density in a given area, gross density, net residential density, floor area ratio and coverage ratio can be calculated depending on different conditions. The emergence of the Geographic Information System technology and Open Data from online sources, such as OpenStreetMap, provides new opportunities for measuring metrics of urban form on the built environment.

Theoretically, a number of researchers acknowledged that the design of urban form on the built environment has been associated with people's social participation and urban vitality. However, few quantitative studies have proven that elements of urban form are important factors in urban vitality, of which, many suggested that a mixture of uses and high density have a significant effect on vitality. Zumelzu concluded that land use mix, block size, plot size and adaptability of building uses have a close association with people's interactions and vitality (Zumelzu, 2019). The land use mix and adaptability of building uses attract various people and encourage social interactions between residents. Similarly, according to Wu, high density and mixed land use has a strong positive correlation with vitality because of stimulation of diverse activities (Wu, 2017). In addition to the level of mixed-use, Long found that vitality increases with intersection density, which is defined as the number of intersections for each square kilometer, accessibility to amenities and transportation that include city center, green space, shopping center, hospital, and a subway station (Long, 2017). Quantitative research on the correlation between street connectivity and urban vitality is relatively new. Street connectivity is often defined as the number of intersections per square kilometer (Long, 2017, Koohsari, 2014). Koohsari illustrated that street connectivity has a positive relationship with urban vitality because a connected street encourages walking for transport and encourages more physical activity (Koohsari, 2014).

Conclusion

The literature review has presented some of the relevant literature on the concept and key factors of urban vitality, the elements and measurements of urban form and qualitative and quantitative analyses on the correlation between urban vitality and urban form. With the development of new techniques and open “big data” on both vitality and urban spatial patterns, there is a new trend on exploring their relationship through a more objective and reliable quantitative analysis based on geographic information system data on urban form, point of interest data, GPS-based mobile phone data, and social media data on vitality. However, the studies by application of social media data are relatively rare. Meanwhile, the metrics of urban form that researchers selected are not fully complete while studying its relationship with vitality. More specifically, some researchers focus on density, a mixture of land use, and urban layout, while others emphasize on how walkability and street connectivity correlate with urban vitality. Therefore, this research will select a full set of urban form variables and conduct an exploratory and quantitative analysis of how these variables influence urban vitality in the following parts which will be described in the following chapter.

Chapter 3: Research Methodology

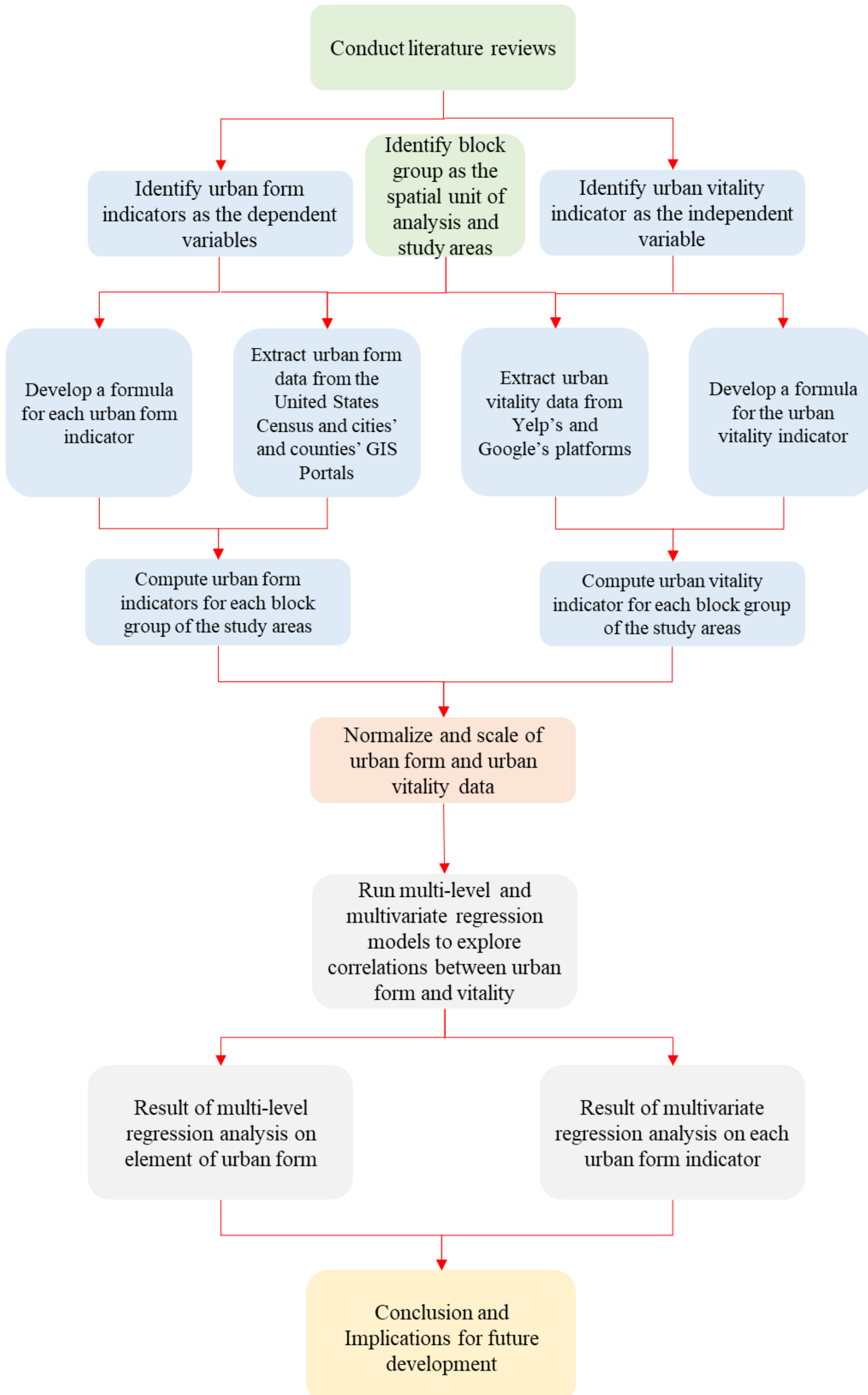
This chapter briefly describes the methods, technique tools, and research framework of this study. The study is mainly focused on exploring how urban form of the built environment influences urban vitality in the downtowns of the three cities in the Texas Triangle. The analysis of the study combines a literature review and a quantitative method, including multi-level and multiple variate regression models over the variables. The study first reviews relevant literature on urban or neighborhood vitality, elements or metrics of urban form and existing studies on their correlations. Most of the literature was found from the University of Texas' online library portal through searching keywords, such as 'form', 'vitality', 'physical aspects' and so on, and others were found from Google Scholar via citations in the previous literature.

Prior to running the multi-level and multivariate regression models, it is necessary to identify the independent and dependent variables for the regression models. In this study, the independent variables are the indicators of urban form. Most of the indicators are extracted from existing literature, including intersection density, average block length, average block size, and socioeconomic variables on population density and employment density. The population density and employment density are control variables in the study. In addition to these, additional indicators of urban form are added, including activity density, sidewalk density index and connected node ratio. The dependent variable in the study is the total number of Yelp reviews. The detail on the definitions, rationale and measurement of variables will be discussed in the next chapter.

Moreover, the study areas and spatial unit of the study are important as well. The downtown areas of the City of Austin, the City of Dallas and the City of San Antonio are identified as the study area and block groups are defined as the spatial unit of the study. For doing research on downtown areas of the three cities in the Texas Triangle, block groups are the most appropriate unit to study. The census tract and census block are not applicable in the study because of various reasons. First, there is a limited size of census tracts in downtown areas of the three cities, providing insufficient sample sizes for the research. In total, the study area only has 15 census tracts in the three cities. Second, it was impossible to generate the number of population and employment on census blocks and this spatial unit was too small to represent and compare varying physical form among each other. In the end, the block group is chosen as the spatial unit for the study. The study area of the City of Austin has 6 block groups, the study area of the City of Dallas has 8 block groups, and the study area of San Antonio has 9 block groups. Therefore, the study has a total of 23 block groups.

Finally, the study calculates indicators of urban form and urban vitality in each block group using Geographic Information System and TransCAD functions based on data extracted from the United States Census, along with the cities'- and counties'- GIS portals and Yelp's platform. Then multi-level and multivariate stepwise regression models are run via SPSS Statistics to analyze the effect of each urban form indicator in urban vitality.

Figure 1: Research Framework



Chapter 4: Research Design & Analytical Process

The study analyzes the relationships between urban form of the built environment and urban vitality in downtown areas of the City of Austin, the City of Dallas and the City of San Antonio from a perspective of how urban vitality is affected by an urban form of the built environment. The research follows a four-step approach, including site selection, independent variables identification, data collection and calculation, dependent variable identification, data collection and calculation, and multi-level and multivariate regression modeling.

Site Selection

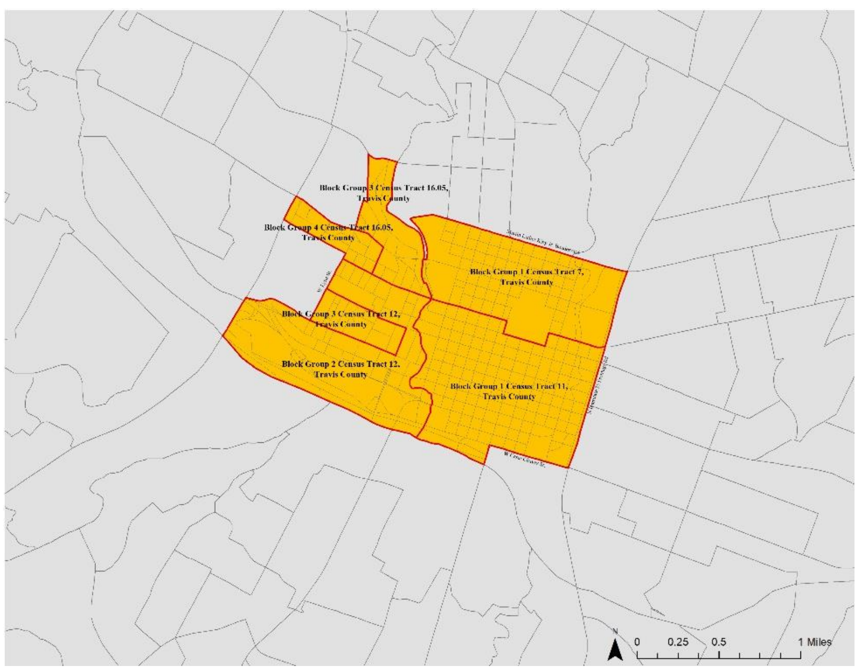
This study has selected downtown areas within three main cities in the Texas Triangle, including Austin, Dallas, and San Antonio, because the Texas Triangle is one of the most growing and flourishing megaregions in the United States. It is estimated that 35 million people will live in the Texas Triangle by 2050 (America 2015). The three cities also have a great diversity in the spatial form of their built environment. To make the selected areas more reasonable, the study follows three principals to select the areas: (1) the selected areas should locate in downtown areas of three major cities within the same megaregion; (2) the selected areas should have approximately the same amount of total employment and residents associated with block groups within the City of Austin, the City of Dallas and the City of San Antonio; (3) the selected areas should include multiple shapes, lengths and sizes of blocks, different street layout, and a mixture of land uses, including commercial, cultural, residential, institutional, and entertainment uses and so on.

Table 1: Population, Employment and Area Estimation for Study Area

City	Block Group	Population	Employment	Area (square miles)
Austin, Travis County	Block Group 3, Census Tract 16.05	627	438	0.183
	Block Group 4, Census Tract 16.05	867	665	0.100
	Block Group 1, Census Tract 7	1121	718	0.624
	Block Group 3, Census Tract 12	809	499	0.108
	Block Group 1, Census Tract 11	4166	2979	0.789
	Block Group 2, Census Tract 12	3139	2344	0.595
Total		**10729**	**7643**	**2.399**
Dallas, Dallas County	Block Group 1, Census Tract 31.01	1887	1434	0.182
	Block Group 2, Census Tract 17.03	1442	1311	0.077
	Block Group 1, Census Tract 21	828	542	0.225
	Block Group 2, Census Tract 17.04	373	300	0.069
	Block Group 1, Census Tract 17.01	494	383	0.236
	Block Group 3, Census Tract 204	2836	2086	0.642
	Block Group 2, Census Tract 21	905	828	0.156
	Block Group 2, Census Tract 31.01	1544	1301	0.182
Total		**10309**	**8185**	**1.769**
San Antonio, Bexar County	Block Group 1, Census Tract 1103	1582	729	0.154
	Block Group 1, Census Tract 1501	1522	866	0.253
	Block Group 1, Census Tract 1101	500	282	0.699
	Block Group 2, Census Tract 1101	681	363	0.259
	Block Group 2, Census Tract 1103	1084	475	0.146
	Block Group 2, Census Tract 1921	1114	656	0.624
	Block Group 1, Census Tract 1401	625	361	0.119
	Block Group 3, Census Tract 1101	2285	1046	0.406
	Block Group 1, Census Tract 1921	1223	718	0.496
Total		**10616**	**5496**	**3.156**

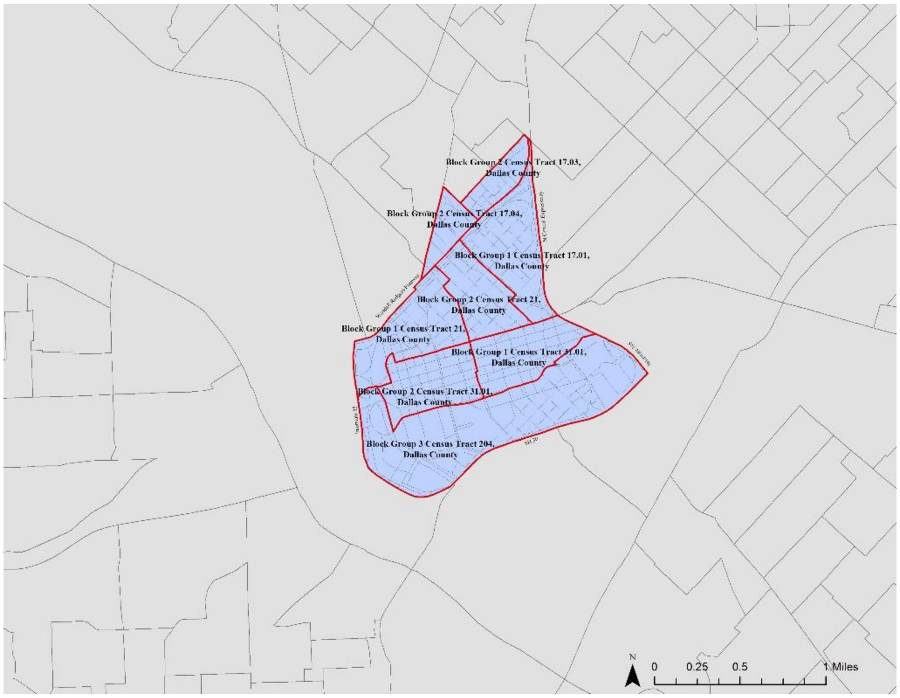
Based on the principles, the study area for the City of Austin contains six block groups showing in Figure 2. It has approximately 10,700 people living in the area, most of which live in block group 1 census tract 11, block group 2, census tract 12 and block group 1 census tract 7, located in the southern and southeast part of the area. Approximately 7,600 people work in the area. Similarly, the jobs are also concentrated in these three block groups. In terms of the size of the block groups, Figure 2 shows that these three block groups are much larger than the other three block groups. The area is shaped with traditional grids made up of blocks grouped as squares or rectangles. The approximate borders of the area include Martin Luther King Jr. Boulevard to the north, W Cesar Chavez Street to the south, West Lynn Street to the west, and N Interstate 35 Frontage Road to the east. The study area has a variety of uses, including the Texas Capitol on the north, the Waterloo Neighborhood Park two blocks east to the Texas Capitol, the Austin Central Library on the corner of West Avenue and W Cesar Chavez Street, the Austin Convention Center on the southeast corner. Most of the commercial and mixed-use development are located on the south and southeast part of the area.

Figure 2: City of Austin Block Group Study Selection



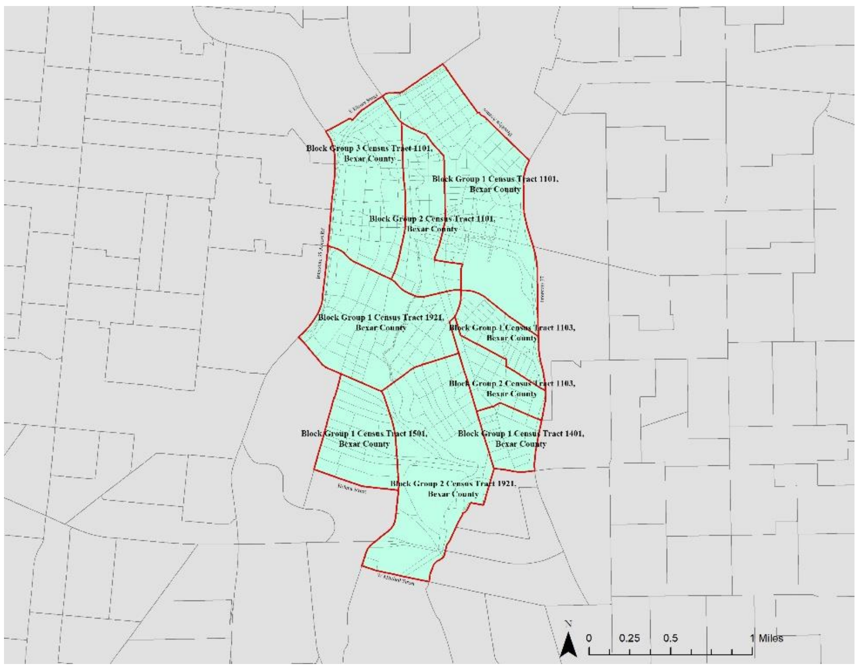
The study area for the City of Dallas contains eight block groups shown in Figure 3 below. The selected block groups have a population of approximately 10,300 and employs around 8,200 people. Table 1 shows the range of population and employment in block groups in the City of Dallas is smaller than that in block groups in the City of Austin. Most of the population and employment are situated in the south of the area as well. Besides block group 3 census tract 204, block group 2 census tract 17.03 and block group 2 census tract 17.04, the size of the rest block groups is relatively similar. The approximate borders of the study area for the City of Dallas include Woodall Rodgers Freeway to the north, SH 30 to the south, Interstate 35 to the west, N Central Expressway to the northeast and Interstate 345 to the east. The irregularly shaped area has various sizes of blocks that have different sharps. Most of the commercial uses are located in the center of area, including Kay Bailey Hutchison Convention Center Dallas located in the southwest corner and Dallas City Hall is located on the east side of the Convention Center. Most of the cultural developments are situated in the northeast, including the Dallas Museum of Art, Crow Museum of Asian Art and Nasher Sculpture Center.

Figure 3: City of Dallas Block Group Study Selection



The study area of the City of San Antonio includes nine block groups shown in Figure 4. It has approximately 10,600 residents and 5,500 employees. Compared to the City of Austin and the City of Dallas, the population and employment are more evenly distributed within the selected block groups of the city of San Antonio. The majority of the small-sized block groups are located in the southeast and southwest portion of the study area. It is surrounded by Interstate 35 Access Road on the west, W Mitchell Street and Helena Street on the south, Interstate 37 on the east, E Elmira Street on the north and Broadway Street, and Brooklyn Avenue on the northeast. The shapes of the block groups are irregular. The area is dominated by a variety of blocks that have a rich diversity of sizes and shapes. Henry B. Gonzalez Convention Center, Tower of the Americas a highrise observation tower-restaurant, the Social Security Office, the Institute of Texan Cultures and Alamo dome which is a multi-purpose stadium are all located in the southeast. Most commercial developments close to the River Walk are located in the center.

Figure 4: City of San Antonio Block Group Study Selection



There are limitations to the chosen study areas. As for doing research on the Texas Triangle, the City of Houston is not selected due to its significantly smaller number of employees in downtown areas with the number of 1800 compared to the other three cities. Therefore, from the perspective of employment, the other three cities have a similar centralized development pattern, while people in San Antonio are more likely to work outside of the downtown. Another limitation to the site selection is the fact that only the downtown areas of the three cities are being studied and therefore the number of population and employees are relatively small. Meanwhile, in the study, the total number of block groups are limited as the sample size is slightly small. Nevertheless, the study on the 23 block groups of the three major cities plays a fundamental effect on exploring the correlations between urban form and urban vitality.

Urban Form Indicators (Independent Variables) Identification

Previous qualitative studies illustrated that the role of urban form of the built environment has an influence on the urban vitality in a city. Before starting a quantitative analysis of urban form, it is important to define specific indicators of it. However, there is a lack of a full set of indicators and quantitative measurements that obstructed the representation of an urban form of the built environment. To better understand and analyze the urban form of the built environment of the three selected study areas, this study defined six indicators of urban form of the urban environment, along with two socioeconomic variables with a hypothesized close association to urban vitality. These indicators were categorized into two categories, including density and design. Three elements of urban form, which include design, urban layout, and street layout as indicators of urban form, are included in these categories. Density will include three indicators, including population density, employment density and activity density. The urban layout design contains average block length and average block size. The street layout design includes three indicators, including sidewalk density index, intersection density, and connected node ratio.

Density is one of the most important urban form indicators affecting urban vitality. The hypothesis on density is that higher densities encourage more population activities. Since the study focuses on the correlation between urban form of the built environment and urban vitality in relation to population activity, it is more appropriate to calculate and compare *population density, employment density and activity density* among block groups. One single density indicator cannot accurately measure the density of the study areas and therefore the three density indicators were selected to provide a more robust and complete measurement of the density (Jenks and Dempsy, 2005). Population density represents how densely populated a block group is. It is calculated by dividing the number of people in each block group by the total area of each block group. Similarly, employment density expresses how densely employed a block group is. It is calculated by dividing the number of jobs in each bock group by the total area of a block group. The larger the value of the population density and employment density, the denser populated and employed the block group. Moreover, in this study, activity was identified as specific uses of land in relation to land use mix, including commercial, retail, government services, educational, institutions, meeting, cultural services and parks, because these uses intentionally have physical activities that people are more intended to participate in. Therefore, activity density is significant in assessing the nature and attractiveness of urban form. In other words, activity density quantifies the density of physical activities in each block group among three cities. It is computed by dividing the total area of the specific uses of land in each block group by the total area of each block group. The smaller the value of the activity density, the fewer physical activities the block group has.

In addition to the densities, many researchers proved that a successful urban space has a good city form design. When defining whether a city form design is good or not, street life is an essential indicator to examine (Jacobs, 1961), and therefore it is important to calculate *sidewalk density index, intersection density and connected node ratio* among block groups. Sidewalk density index quantifies how walkable a block group is within the three study areas. However, there is a limitation to compare the traditional measurement of the total length of sidewalks. It is obvious that the value of the total length of sidewalks is varying with diverse sizes of the block groups. For instance, the more area of the block groups is, the longer the sidewalk is. Therefore, comparing the total length of sidewalks of each block group is not reasonable because of the multi-sized block groups in the study. Instead, a sidewalk density index is more appropriate in this research. It was calculated by dividing the sum of the total length of sidewalks by area of each block group. In addition to pedestrian flow, city form design on street layout for traffic flow and connectivity should also be considered. Both intersection density and connected node ratio represent connectivity of the street network. Intersection density, calculated by dividing the total number of intersections by area of each block group, quantifies how many intersections per block group exist, while connected node ratio measures the ratio between the numbers of non-dead street intersections, also called the real intersection, and the numbers of total street intersections. Obviously, the figure 4 shows the cul-de-sac is a node that only has a one-way connection while the real intersection is a node that has more than two-way connections. The design on the urban layout that includes *average block length* and *average block size* has a close association with population activities.

Hypothetically, the shorter the block length or the smaller the block size will produce more intersections within a block group, therefore providing a larger number of routes for populations (Ye, 2017). Thus, people are more likely to walk places where a shorter block length and a smaller block size exist. The average block length is calculated by the total distance of streets divided by the total number of nodes. However, the average block size is computed by the total area of the block in each block group dividing by the number of the census blocks in that block group. Table 2 summarizes the indicators listed above and includes information on the calculation and data sources of the variables.

Table 2: Summary of Socioeconomic & Urban Form Indicators

Type	Indicator	Calculation	Interpretation	Data Source
Density	Population Density	Number of people in the block group/total area of a block group	Density of population	U.S. Census Bureau, GIS Database
	Employment Density	Number of jobs in the block group/total area of a block group	Density of employment	U.S. Census Bureau, GIS Database
	Activity Density	Area of specific uses of the land/total area of a block group	Density of physical activities	GIS Database
City Form Design	Walkable Sidewalk Length Ratio	Sum of sidewalks in a block group/total area of a block group	Level of walkability	GIS Database
	Intersection Density	Number of street intersections/total area of a block group	Level of connectivity	GIS Database
	Connected Node Ratio	Number of non-dead street intersections/total number of street intersections	Level of connectivity	GIS Database
	Average Block length	Total distance of streets/number of nodes	Urban Layout	GIS Database
	Average Block Size	Total area of the block in a block group/number of blocks	Urban Layout	GIS Database

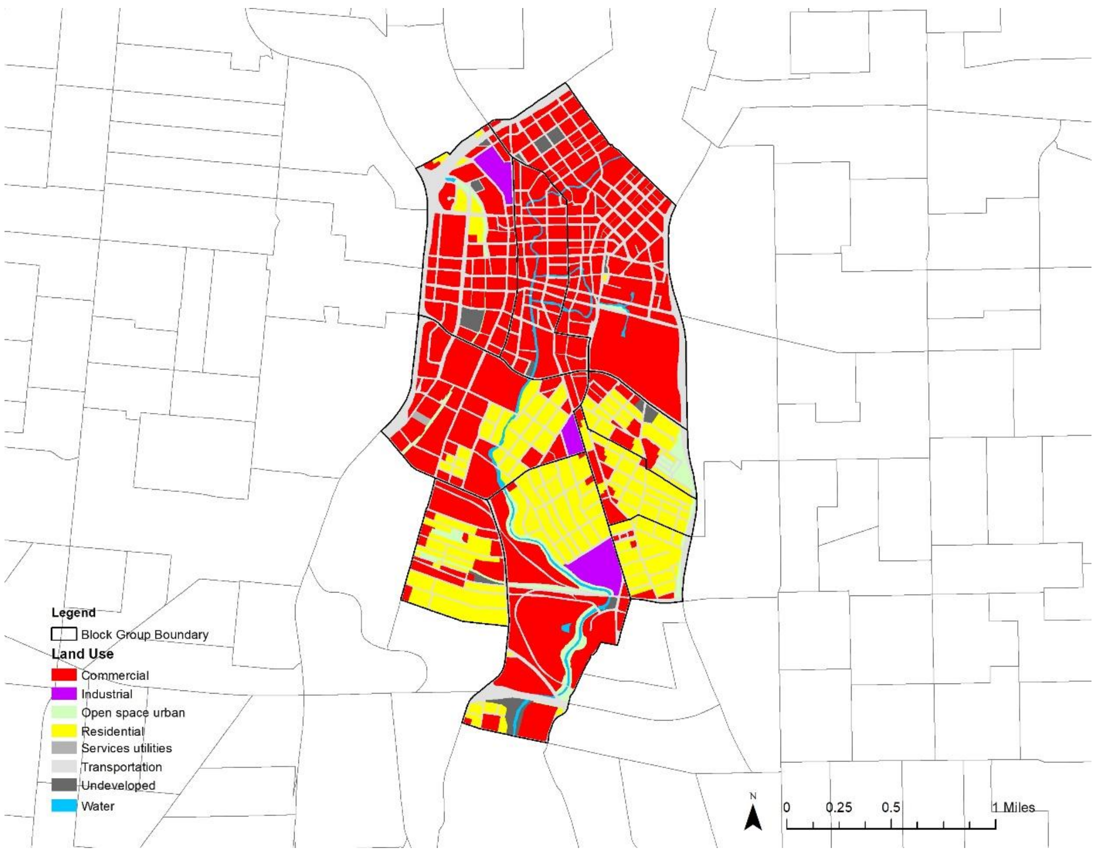
Data Sources on Independent Variables

After defining indicators of urban form of the built environment within the three main cities, the next step is to collect the data, process the data and use the data to subsequently calculate results for each urban form indicator for each block group. In the study, the data on population and employment can be drawn from the United States Census Bureau, 2012-2016 American Community Survey 5-year Estimates. The raw data is obtained from American Factfinder and contains population and employment data for each block group within Travis County, Dallas County and Bexar County. The rest of spatial data, including land use, street segment, sidewalk, and census blocks are collected from the cities'- and counties'- geographic information system portals. These datasets are used to calculate each of the independent variables. Specifically, the land use data includes categories of land uses and is used to calculate activity density. The sidewalk vector data, which provides detailed information on the length of each sidewalk, is used to calculate a sidewalk density index in each block group. In addition, street segment data is needed to calculate average block length, intersection density and connected node ratio. Lastly, census blocks data is used to calculate the average block size in each block group.

Data Imputation and Processing Procedure

The GIS data on the city of Dallas's sidewalks is missing. In order to collect the most accurate data on the sidewalk, the distance of sidewalks within each block group in the study area of the City of Dallas were manually measured by using the Google Earth measuring tool in order to calculate a sidewalks density index. Unfortunately, the land use data on the City of San Antonio is only categorized into commercial, industrial, open space urban, residential, service utilities, transportation, undeveloped and water. Meanwhile, the dataset is not intended to provide information at the property level. As it is shown in Figure 5 below, land uses for the three block groups in the north portion within in the study area are mostly commercial uses.

Figure5: Land Use Map within the Study Area for the City of San Antonio



Thus, this dataset is completely misleading for comparing activity density among block groups in the City of San Antonio. Without the precise land use information and map from the city's portal, the amount of commercial, retail, government services, educational, institutions, meeting, cultural services and parks uses in the city of San

Antonio cannot be predicted. Moreover, it was inapplicable to impute the value by the traditional data imputation methods, like mean imputation, 0 imputation and median imputation because there was no direct relationship on special land uses among each block group in the three cities. The number of land uses is also randomly distributed. To get the land use data of the selected area in the City of San Antonio, information was used from Google Maps, including buildings' names and their representative colors (e.g. light orange represents commercial and green represents parks) to measure areas of the specific land uses in the block groups of the City of San Antonio.

In the processing procedure, the Geoprocessing function provided by ArcGIS and Join functions provided by TransCAD are used to calculate the level of connectivity. In order to calculate Intersection Density and Connected Node Ratio, it is necessary to calculate the number of cul-de-sacs and real intersections in each block group within the three cities. The study performed the following steps to calculate them:

- Utilized Intersect function to extract street segment shapefile within the study areas from the cities' street files in ArcGIS
- Utilized Select by Attribute function to exclude freeway and ramp
- Input Street segment shapefile and exported it to standard geographic file, which includes both line and point attributes in TransCAD
- Extracted the IDs of the nodes of each street, which include both From ID and To ID as origin and destination separately from the standard geographic file of streets
- Joined nodes to the origin and destination of streets through the same IDs

- Defined the type of intersections categorized by number of links (line attributes) connected to each node
- Summarized of FromNode and ToNode to get how many ways the nodes connect and then identify 1-way, 3-way, and 4-way intersections, ignoring nodes with two links
- Filtered the results and removed those obviously wrong 1-way intersections due to the disconnected street from ArcGIS
- Simplified and finalized the results with the number of cul-de-sac (1-way intersection) and real intersections (>=3-way intersection) in each block group

The following figures demonstrate that the location and number of cul-de-sac and real intersections in each block group of three study areas in the City of Austin, the City of Dallas and the City of San Antonio shown in Figure 6, 7, and 8. In the study area of Austin, out of 562 intersections, 451 intersections are real intersections. Depending on the large size of the block groups, the majority of the real intersections are located in block group 1 census tract 7, block group 1 census tract 11 and block group 2 census tract 2. However, block group 3 census tract 12 has the least number of real intersections, but it only has two dead-end nodes. Out of 616 intersections in the study area of the City of Dallas, 508 are real intersections. Block group 3 census tract 204 has the highest number of real intersections, while block group 2 census tract 17.03 and block group 2 census tract 17.04 have the lowest number of real intersections with 33 and 27 respectively. The rest of the block groups have a relatively similar number of intersections. Compared to the City of Austin and Dallas, there are three levels of the number of intersections in the study area of the city of San Antonio. Block group 1 census tract 1101 has the most real intersections with 175 out of 197 total intersections. The second level includes block group 1 census tract 1921, block group 3 census tract 1101, the block group 2 census tract 1101 and block group 2 census tract 1921, which all have approximately 70 real intersections. In addition, the number of real intersections for the remaining four block groups is around 40.

Figure 6: Intersections within the Study Area for the City of Austin

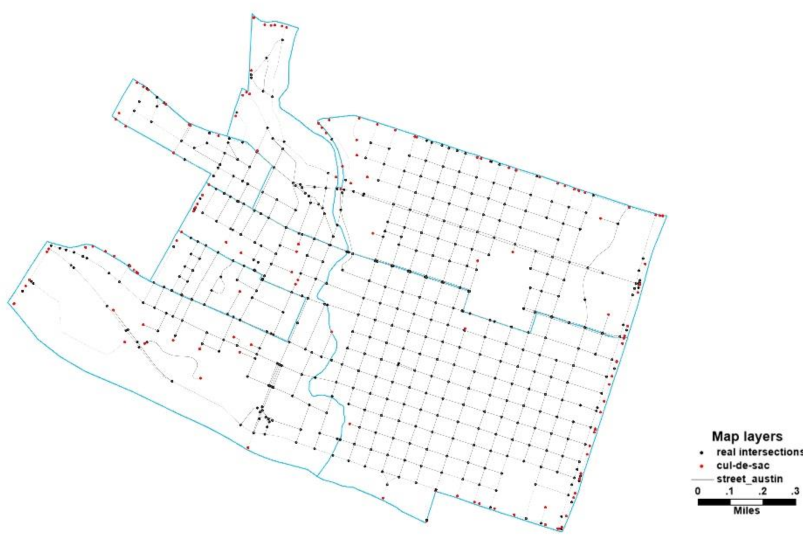


Figure 7: Intersections within the Study Area for the City of Dallas

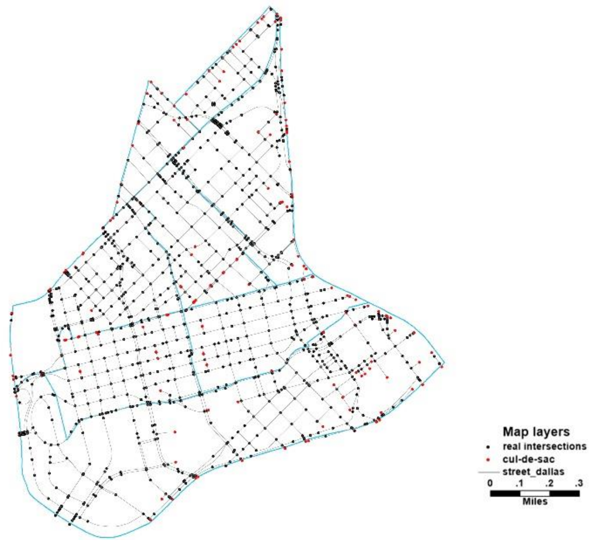
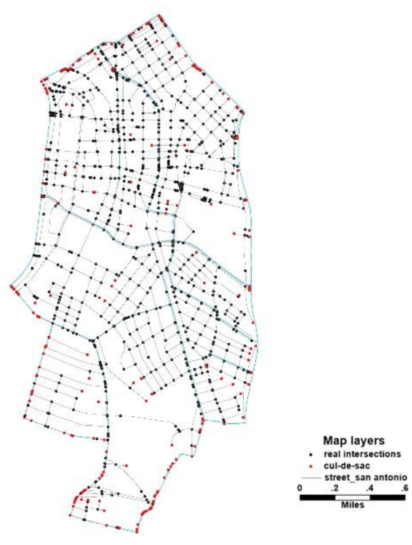


Figure 8: Intersections within the Study Area for the City of San Antonio



The remaining independent variables that include population density, employment density, activity density, sidewalk density index, average block length and average block size, are calculated by utilizing ArcGIS's geoprocessing function and Excel's formula function individually.

- Utilized intersection function to extract all spatial shapefile data within our focus area from the cities' files in ArcGIS
- Recorded area of each block group, number of populations, number of employments, length of streets, length of sidewalks, area of specific uses of land and number of census blocks in each block group through Selection by Attribute and Statistic function in ArcGIS
- Calculated and finalized the result on each urban form indicator by using their formulas in Excel sheet

Preliminary Result on Urban Form Indicators

The table 3 summarizes the preliminary results of two socioeconomic and six urban form indicators as independent variables by each block group. The results show that the 23 block groups of the three cities identified in the study have a diversity of population and employment density. Most block groups have population density ranging from 3,000 to 9,000 per square mile and employment density ranging from 2,000 to 8,000 per square mile, but there are still block groups with extremely high and low density. For example, block group 2 census tract 17.03 in Dallas has both the highest population density and employment density. With regard to activity density, the percentage of activity density in the 23 block groups varies from 4.50 to 76.05. This wide range of activity density helped to develop a relatively comprehensive analysis of its relation to urban vitality. Most block groups in the City of Dallas tend to have a higher percentage of activity density than the block groups in the City of Austin and most block group in the City of San Antonio, except block group 1 census tract 1101 and the block group 2 census tract 1101.

In addition to activity density, different values, representing a diversity of spatial form among 23 block groups, are also reflected in the rest of the urban form variables. With regard to the two urban layout variables, most block groups in the City of Austin have a larger average block size than the block groups in the other two cities, which implies that vitality tends to be weak in those larger block groups. The table also shows that the selected block groups in the City of Dallas have an average block size of less than 150000. Similarly, compared to the study area in Austin and San Antonio, most block groups in Dallas have much less average block length than those in Austin and San Antonio. With respect to the three street layout variables, based on the higher values on sidewalk density index and connected node ratio, the block groups in the City of Dallas provide better-connected streets and walkable sidewalks, which intends to promote more pedestrian activities.

However, as it is shown in table 3, the preliminary results on independent variables are highly varying in magnitudes, units and range. For instance, the results include both percentages and numbers. Meanwhile, the results are varying from large magnitudes on average block size, population density, and employment density to small magnitudes on the sidewalk density index. In order to avoid one variable like average block size weighing more heavily over other variables, it is important to normalize the independent variables that transformed the original values into new values which had the relatively same range. Prior to normalization, for the percentage on activity density and connected node ratio, the values were multiplied by 100 into regular numbers. Then, as Ye and Venerandi did in their research (Ye, 2017 & Venerandi, 2017), the independent variables' results were normalized through log transformation and subsequently scaled the results by z score standardization. The formula for the transformation and z score standardization is shown below:

The log transformation can be done by the formula: $y_i = log10\ (x_i)$.

The z score normalization can be done by the formula: $y_i = (xi - \mu) / \sigma$,

Where μ *is the mean of* x *and* σ *is the standard deviation of* x.

Table 3: Independent Variables' Results of Study Area by Block Groups

City	Block Group	Population Density (per Square Miles)	Employment Density (per Square Miles)	Activity Density (in relation to land use mix)	Sidewalks Density Index	Intersection Density (per Square Miles)	Connected Node Ratio	Average Block Length (feet)	Average Block Size (sf)
Austin, Travis County	Block Group 3, Census Tract 16.05	3418	2388	45.70	39.77	207	74.51	559	300823
	Block Group 4, Census Tract 16.05	8664	6646	4.90	42.06	270	62.79	481	199260
	Block Group 1, Census Tract 7	1798	1151	45.30	36.25	172	74.83	754	235064
	Block Group 3, Census Tract 12	7513	4634	12.49	42.72	251	93.10	583	250167
	Block Group 1, Census Tract 11	5279	3775	35.67	44.79	195	88.00	750	149674
	Block Group 2, Census Tract 12	5273	3938	50.46	30.34	165	80.99	798	263414
Dallas, Dallas County	Block Group 1, Census Tract 31.01	10341	7858	58.06	52.21	279	80.95	634	110585
	Block Group 2, Census Tract 17.03	18671	16975	29.48	45.10	427	80.49	383	107183
	Block Group 1, Census Tract 21	3685	2412	44.99	44.64	338	82.61	666	86653
	Block Group 2, Census Tract 17.04	5407	4349	73.12	50.22	391	67.50	400	105877
	Block Group 1, Census Tract 17.01	2094	1623	74.21	52.75	301	78.02	770	94829
	Block Group 3, Census Tract 204	4418	3250	48.95	42.42	212	82.93	963	147667
	Block Group 2, Census Tract 21	5817	5322	71.03	46.32	296	92.00	652	93969
	Block Group 2, Census Tract 31.01	8487	7151	67.98	61.23	374	90.67	558	125880
San Antonio, Bexar County	Block Group 1, Census Tract 1103	10272	4734	11.86	44.46	247	88.37	683	93336
	Block Group 1, Census Tract 1501	6022	3427	9.31	27.26	123	67.39	719	251627
	Block Group 1, Census Tract 1101	715	403	60.63	31.33	250	88.83	734	126543
	Block Group 2, Census Tract 1101	2627	1400	76.05	42.15	285	93.67	719	86051
	Block Group 2, Census Tract 1103	7416	3249	11.63	50.96	328	90.57	537	116435
	Block Group 2, Census Tract 1921	1785	1051	15.91	20.48	111	77.53	785	228933
	Block Group 1, Census Tract 1401	5264	3040	4.50	41.60	286	79.07	488	143914
	Block Group 3, Census Tract 1101	5628	2576	57.92	32.74	239	80.83	821	116686
	Block Group 1, Census Tract 1921	2464	1447	31.14	30.21	153	84.44	877	177373

Urban Vitality Indicator Identification and Data Collection

Social media platforms have been changing our lives for a long time. Nowadays, most people are sharing their life, stories and thoughts on a social media platform. For instance, many people are likely to post pictures and comments while they are eating at a lovely restaurant. Yelp is one of the most popular social media platforms where users can follow and share their similar interests with each other. Given that a large amount of people is using Yelp, it is a valuable resource to help people to find their favorite meeting places and write reviews. The study used the number of Yelp reviews on each block group as an urban vitality indicator in this study. In order to generate the most representative indicator of a city vitality from Yelp, the most recent reviews on a whole set of categories of businesses were totaled, including restaurants, public services & government, education, arts & entertainment, parks & recreation, hotels & travel, health & medical and religious organizations. This complete category of businesses almost contains every place where people are intended to meet. The Yelp reviews that were collected in the study covers the time period from January 1st, 2019 to June 30th, 2019. These recent reviews minimize the possibility of misleading information resulted from the old businesses that have been closed for a long time. However, the limitation on Yelp reviews was that there were many fewer reviews on parks and recreation than other categories, leading to obviously inaccurate results in the study. To reduce its effect on the study's implications, Google reviews on parks and recreation from January 1st, 2019 to June 30th, 2019 will be included in the Yelp review data. The dependent variable in the study is the total number of Yelp reviews and Google reviews on parks & recreation. The equation for calculating it is shown below:

Total # of Yelp reviews = # of Yelp reviews on restaurants + # of Yelp reviews on public services & government + # of Yelp reviews on education + # of Yelp reviews on arts & entertainment + # of Yelp reviews on parks & recreation + # of Yelp reviews on hotels &

travel, # of Yelp reviews on health & medical +# of Yelp reviews on religious organization + # of google reviews on parks & recreation.

In calculating and comparing the number of reviews of each block group among three study areas in the City of Austin, the City of Dallas and the City of San Antonio, the study first generated number of reviews for each category with each block group based on their geographic location. The number of reviews on parks & recreation of each block group on Google were added to the dataset, and then subsequently summed for every block group. The urban vitality data is obtained from Yelp's and Google's platform. Both Yelp's and Google's platforms can search places based on maps, which is useful for the study to generate reviews block group by block group. A total of 22,308 reviews are collected in the study.

Preliminary Result on Socioeconomic & Urban Vitality Indicator

Table 4 illustrates the preliminary results of the urban vitality indicator as a dependent variable. Like independent variables, the preliminary results on the number of Yelp reviews per block groups have different magnitudes. In other words, the reviews on businesses are varying between 1 in Block Group 3, Census Tract 1101, San Antonio, Bexar County and 6, 038 in Block Group 1, Census Tract 11, Austin, Travis County. Therefore, it was necessary to normalize the preliminary result prior to running regression models for the correlation between urban form indicators and urban vitality indicator. The study normalized the data with logarithmic transformation in the same manner as the independent variables. Furthermore, it was obvious that there were outliers in the dependent variable. A z-score normalization was used to scale the feature of the dependent variable.

Table 4: Urban Vitality Indicator's Results of Study Area by Block Groups

City	Block Group	**Yelp's Reviews**
Austin, Travis County	Block Group 3, Census Tract 16.05	258
	Block Group 4, Census Tract 16.05	18
	Block Group 1, Census Tract 7	488
	Block Group 3, Census Tract 12	167
	Block Group 1, Census Tract 11	6386
	Block Group 2, Census Tract 12	1022
Dallas, Dallas County	Block Group 1, Census Tract 31.01	1287
	Block Group 2, Census Tract 17.03	93
	Block Group 1, Census Tract 21	534
	Block Group 2, Census Tract 17.04	341
	Block Group 1, Census Tract 17.01	317
	Block Group 3, Census Tract 204	541
	Block Group 2, Census Tract 21	2341
	Block Group 2, Census Tract 31.01	1197
San Antonio, Bexar County	Block Group 1, Census Tract 1103	105
	Block Group 1, Census Tract 1501	22
	Block Group 1, Census Tract 1101	3742
	Block Group 2, Census Tract 1101	1439
	Block Group 2, Census Tract 1103	141
	Block Group 2, Census Tract 1921	353
	Block Group 1, Census Tract 1401	1
	Block Group 3, Census Tract 1101	528
	Block Group 1, Census Tract 1921	987

As it is shown in Table 4, in the City of Austin, block group 1 census tract 11 in the south of the study area has many more reviews than other block groups. Thus, it is suggested that the south and southeast parts of the area have a dramatically stronger urban vitality than the north and the west of the area. Moreover, out of the six block groups in Austin, block group 4 census tract 16.05 has the least number of Yelp reviews. As expected, most of the reviews are generated in the block groups near the center of study area in the City of Dallas that include block group 2 census tract 21, block group 2 census tract 31.01, block group 1 census tract 31.01 and block group 1 census tract 21, indicating that the urban vitality is mostly concentrated in the center of the study area in the City of Dallas. However, in the study area of the City of San Antonio, the block groups in the north of the area have a higher number of Yelp reviews than others. Compared to the north of the area, the block groups in the southeast have many fewer reviews. It is especially important to note that block group 1 census tract 1401 only has 1 review because of a lack of activity places in this block group.

Summary of Preliminary Result and Normalization

Table 5: Summary of Descriptive Statistics of Variables

No.	Variables	Min	Max	Mean	Standard Deviation
1	Population Density	715	18671	5785	3892.35
2	Employment Density	403	16975	4035	3445.69
3	Activity Density	4.50	76.05	40.93	24.51
4	Sidewalk Density Index	20.48	61.23	41.39	9.47
5	Intersection Density	111	427	257	83.43
6	Connected Node Ratio	62.79	93.67	81.74	8.49
7	Average Block Length	383	963	666	149.07
8	Average Block Size	86051	300823	157041	66645.13
9	Yelp Reviews	1	6386	970	1467.47

The table 5 above shows the summary of descriptive statistics of the nine variables. The dependent variable of Yelp reviews ranges from 1 to 6386 per block groups. The data on Yelp reviews are long-tailed distributed, which show that there are more block groups with fewer reviews than block groups with more reviews. Similar to the Yelp reviews, the socioeconomic independent variables, employment density and the urban form independent variable, average block size are also long-tailed distributed with a larger number of low values occurring across the block groups. While, based on the statistics on population density, it is right-skewed normally distributed. The rest of the variables seem to be normally distributed. To justify the assumption, a Kolmogorov-Smirnov test was conducted via the export function under Descriptive Statistics in SPSS Statistics to check the normality of each variable. The summary statistics are represented in Table 6. The Kolmogorov-Smirnov test is one of the well-known tests of normality in statistical analysis. The result shows that the employment density, average block size and Yelp reviews are not normally distributed with a p-value greater than 0.05.

Table 6: Summary Statistics of the Kolmogorov-Smirnov Test on Variables

Tests of Normality			
	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
Population Density	0.171	23	0.078
Employment Density	0.202	23	0.016
Activity Density	0.151	23	0.191
Sidewalk Density Index	0.161	23	0.126
Intersection Density	0.079	23	.200*
Connected Node Ratio	0.117	23	.200*
Average Block Length	0.118	23	.200*
Average Block Size	0.198	23	0.020
Yelp's Review	0.267	23	0.000

To prepare the data for the regression models, as mentioned earlier, a Logarithmic transformation was used to normalize variables that include population density, employment density, average block size and Yelp reviews. The results of this transformation are summarized in Table 7 below.

Table 7: Results of Logarithmic Transformation

City	Block Group	LOG Population Density (per Square Miles)	LOG Employment Density (per Square Miles)	Activity Density (in relation to land use mix)	Sidewalks Density Index	Intersection Density (per Square Miles)	Connected Node Ratio	Average Block Length (feet)	LOG Average Block Size (sf)	LOG Yelp's Reviews
Austin, Travis County	Block Group 3, Census Tract 16.05	3.53	3.38	45.70	39.77	207.15	74.51	559	5.48	2.41
	Block Group 4, Census Tract 16.05	3.94	3.82	4.90	42.06	269.83	62.79	481	5.30	1.26
	Block Group 1, Census Tract 7	3.25	3.06	45.30	36.25	171.58	74.83	754	5.37	2.69
	Block Group 3, Census Tract 12	3.88	3.67	12.49	42.72	250.74	93.10	583	5.40	2.22
	Block Group 1, Census Tract 11	3.72	3.58	35.67	44.79	195.13	88.00	750	5.18	3.81
	Block Group 2, Census Tract 12	3.72	3.60	50.46	30.34	164.63	80.99	798	5.42	3.01
Dallas, Dallas County	Block Group 1, Census Tract 31.01	4.01	3.90	58.06	52.21	279.49	80.95	634	5.04	3.11
	Block Group 2, Census Tract 17.03	4.27	4.23	29.48	45.10	427.29	80.49	383	5.03	1.97
	Block Group 1, Census Tract 21	3.57	3.38	44.99	44.64	338.21	82.61	666	4.94	2.73
	Block Group 2, Census Tract 17.04	3.73	3.64	73.12	50.22	391.42	67.50	400	5.02	2.53
	Block Group 1, Census Tract 17.01	3.32	3.21	74.21	52.75	300.93	78.02	770	4.98	2.50
	Block Group 3, Census Tract 204	3.65	3.51	48.95	42.42	211.88	82.93	963	5.17	2.73
	Block Group 2, Census Tract 21	3.76	3.73	71.03	46.32	295.66	92.00	652	4.97	3.37
	Block Group 2, Census Tract 31.01	3.93	3.85	67.98	61.23	373.78	90.67	558	5.10	3.08
San Antonio, Bexar County	Block Group 1, Census Tract 1103	4.01	3.68	11.86	44.46	246.74	88.37	683	4.97	2.02
	Block Group 1, Census Tract 1501	3.78	3.53	9.31	27.26	122.66	67.39	719	5.40	1.34
	Block Group 1, Census Tract 1101	2.85	2.61	60.63	31.33	250.35	88.83	734	5.10	3.57
	Block Group 2, Census Tract 1101	3.42	3.15	76.05	42.15	285.41	93.67	719	4.93	3.16
	Block Group 2, Census Tract 1103	3.87	3.51	11.63	50.96	328.37	90.57	537	5.07	2.15
	Block Group 2, Census Tract 1921	3.25	3.02	15.91	20.48	110.56	77.53	785	5.36	2.55
	Block Group 1, Census Tract 1401	3.72	3.48	4.50	41.60	286.36	79.07	488	5.16	0.00
	Block Group 3, Census Tract 1101	3.75	3.41	57.92	32.74	238.92	80.83	821	5.07	2.72
	Block Group 1, Census Tract 1921	3.39	3.16	31.14	30.21	153.14	84.44	877	5.25	2.99

In order to transform the independent and dependent variables into new values that have the same range, mostly from -2 to 2, it is necessary to scale the variables. The study used z-score standardization for the regression. The results are shown in Table 8. Finally, the z-score values of each independent variable and the dependent variable were fed into multi-level and multivariate linear regression models to identify the correlation between urban form indicators and the urban vitality indicator by measuring the coefficient beta.

Table 8: Results of Z-score Standardization

City	Block Group	Zscore Population Density (per Square Miles)	Zscore Employment Density (per Square Miles)	Zscore Activity Density (in relation to land use mix)	Zscore Sidewalks Density Index	Zscore Intersection Density (per Square Miles)	Zscore Connected Node Ratio	Zscore Average Block Length (feet)	Zscore Average Block Size (sf)	Zscore Yelp's Reviews
Austin, Travis County	Block Group 3, Census Tract 16.05	0.4251	-0.3019	0.1946	-0.1709	-0.5919	-0.8516	0.7159	1.8123	-0.1280
	Block Group 4, Census Tract 16.05	-0.8636	0.9817	-1.4696	0.0712	0.1594	-2.2312	1.2376	0.7901	-1.5146
	Block Group 1, Census Tract 7	1.3154	-1.2165	0.1785	-0.5427	-1.0182	-0.8144	-0.5888	1.2002	0.2040
	Block Group 3, Census Tract 12	-0.6660	0.5296	-1.1602	0.1404	-0.0694	1.3374	0.5522	1.3547	-0.3545
	Block Group 1, Census Tract 11	-0.1770	0.2724	-0.2143	0.3586	-0.7360	0.7366	-0.5655	0.0801	1.5432
	Block Group 2, Census Tract 12	-0.1756	0.3254	0.3891	-1.1672	-1.1016	-0.0885	-0.8882	1.4828	0.5889
Dallas, Dallas County	Block Group 1, Census Tract 31.01	-1.1086	1.1919	0.6990	1.1429	0.2751	-0.0931	0.2124	-0.6710	0.7090
	Block Group 2, Census Tract 17.03	-1.9272	2.1577	-0.4668	0.3916	2.0468	-0.1478	1.8996	-0.7486	-0.6593
	Block Group 1, Census Tract 21	0.3210	-0.2892	0.1657	0.3433	0.9790	0.1019	-0.0004	-1.2762	0.2509
	Block Group 2, Census Tract 17.04	-0.2104	0.4500	1.3137	0.9321	1.6168	-1.6768	1.7795	-0.7790	0.0173
	Block Group 1, Census Tract 17.01	1.1041	-0.7857	1.3581	1.1996	0.5321	-0.4381	-0.6994	-1.0525	-0.0207
	Block Group 3, Census Tract 204	0.0694	0.0847	0.3274	0.1083	-0.5352	0.1393	-1.9908	0.0466	0.2577
	Block Group 2, Census Tract 21	-0.3115	0.7032	1.2284	0.5203	0.4690	1.2075	0.0955	-1.0751	1.0206
	Block Group 2, Census Tract 31.01	-0.8349	1.0737	1.1037	2.0958	1.4054	1.0505	0.7252	-0.3496	0.6712
San Antonio, Bexar County	Block Group 1, Census Tract 1103	-1.0994	0.5563	-1.1860	0.3238	-0.1173	0.7804	-0.1189	-1.0919	-0.5961
	Block Group 1, Census Tract 1501	-0.3596	0.1511	-1.2900	-1.4928	-1.6046	-1.6896	-0.3564	1.3692	-1.4101
	Block Group 1, Census Tract 1101	2.5921	-2.5316	0.8038	-1.0628	-0.0741	0.8346	-0.4567	-0.3365	1.2648
	Block Group 2, Census Tract 1101	0.7900	-0.9713	1.4329	0.0801	0.3461	1.4042	-0.3563	-1.2935	0.7671
	Block Group 2, Census Tract 1103	-0.6479	0.0845	-1.1952	1.0104	0.8610	1.0387	0.8666	-0.5431	-0.4426
	Block Group 2, Census Tract 1921	1.3252	-1.3307	-1.0205	-2.2092	-1.7497	-0.4962	-0.8002	1.1346	0.0353
	Block Group 1, Census Tract 1401	-0.1732	0.0012	-1.4861	0.0216	0.3576	-0.3147	1.1955	-0.0173	-3.0199
	Block Group 3, Census Tract 1101	-0.2658	-0.2065	0.6933	-0.9139	-0.2111	-0.1071	-1.0437	-0.5378	0.2450
	Block Group 1, Census Tract 1921	0.8783	-0.9301	-0.3994	-1.1805	-1.2393	0.3180	-1.4147	0.5014	0.5708

Multi-level and Multivariate Regression Models

To explore the association of urban form indicators with urban vitality, an ordinary least squares based (OLS) multi-level and multivariate regression models were run. The equation of the multivariate linear regression model is as follows:

$$Y_{\textit{YELP REVIEWS}} = b_0 + b_1X_{\textit{POPULATION DENSITY}} + b_2X_{\textit{EMPLOYMENT DENSITY}} + b_3X_{\textit{ACTIVITY DENSITY}} + b_4X_{\textit{INTERSECTION DENSITY}} + b_5X_{\textit{SIDEWALK DENSITY INDEX}} + b_6X_{\textit{CONNECTED NODE RATIO}} + b_7X_{\textit{AVERAGE BLOCK LENGTH}} + b_8X_{\textit{AVERAGE BLOCK SIZE}}$$

A Pearson's Correlation was first used to identify possible multicollinearity between independent variables. Table 9 shows correlations for the independent variable. Most of the absolute values of the correlation coefficients of independent variables are less than 0.70 which means no collinearity in these variables. However, some of the independent variables are highly correlated with each other. The correlation between population density and employment density is 0.966 that indicates the substantial collinearity and the intersection density, and the sidewalk length ratio also have collinearity with a number of -0.782.

Table 9: Summary Statistics of Pearson's Correlation

Correlations										
		Yelp's Review	Population Density	Employment Density	Activity Density	Sidewalk Length Ratio	Intersection Density	Connected Node Ratio	Average Block Length	Average Block Size
Pearson Correlation	Yelp's Review	1.000	-0.336	-0.254	0.683	-0.042	-0.194	0.483	0.465	-0.204
	Population Density	-0.336	1.000	0.966	-0.299	-0.477	0.429	-0.027	-0.535	-0.143
	Employment Density	-0.254	**0.966**	1.000	-0.167	-0.530	0.486	-0.080	-0.567	-0.132
	Activity Density	0.683	-0.299	-0.167	1.000	-0.317	0.270	0.209	0.182	-0.442
	Sidewalk Length Ratio	-0.042	-0.477	-0.530	-0.317	1.000	-0.782	-0.248	0.469	0.567
	Intersection Density	-0.194	0.429	0.486	0.270	**-0.782**	1.000	-0.082	-0.750	-0.602
	Connected Node Ratio	0.483	-0.027	-0.080	0.209	-0.248	-0.082	1.000	0.250	-0.389
	Average Block Length	0.465	-0.535	-0.567	0.182	0.469	-0.750	0.250	1.000	0.142
	Average Block Size	-0.204	-0.143	-0.132	-0.442	0.567	-0.602	-0.389	0.142	1.000

Next, tolerance values, the amount of variability in one independent variable, were applied as a more accurate indicator to verify the collinearity on the independent variables and ascertain which one of them will be dropped for a subsequent linear regression model. Generally, tolerance value less than 0.10 indicates collinearity. Table 10 presents the collinearity statistic on tolerance of each independent variable. The population density, employment density and intersection density have a tolerance value of less than 0.1. The collinearity occurring in the independent variables will result in misrepresenting interpretation in the regression models. Therefore, to solve the problem of it, two independent variables were dropped from the analysis, including employment density and intersection density, as population density contains more information than employment density in the study.

Table 10: Results of Tolerance Values on Independent Variables

Coefficients[a]					
Model		Unstandardized Coefficients		Collinearity Statistics	
		B	Std. Error	Tolerance	VIF
1	(Constant)	0.000	0.137		
	Population Density	-0.546	0.751	**0.035**	28.628
	Employment Density	0.611	0.744	**0.036**	28.066
	Activity Density	0.648	0.239	0.345	2.895
	Sidewalk Length Ratio	-0.013	0.304	0.213	4.699
	Intersection Density	-0.444	0.508	**0.077**	13.071
	Connected Node Ratio	0.322	0.203	0.476	2.100
	Average Block Length	0.001	0.332	0.179	5.595
	Average Block Size	-0.050	0.267	0.276	3.619

a. Dependent Variable: Yelp's Review

Chapter 5: Results

Finalizing the analysis, a multiple level regression models was conducted to identify which element of the three categories of urban form variables (density, urban layout and street layout design) are contributing more to create urban vitality. More specifically, a stepwise regression function was also utilized in SPSS to analyze each independent variable in the model individually and returns only those that allow the model to achieve a 90% level of confidence. This chapter presents the results of the data analysis and the regression models.

Correlations between Elements of Urban Form and Urban Vitality

In model 1, the socioeconomic independent variable, population density is introduced and analyzed the correlation with urban vitality. However, according to the significance of the number of 0.117 that is higher than 0.1 shown in Table 11 below, there is no significant evidence that population density has an impact on urban vitality in the study. The result shows the different findings with previous research, such as Yue's finding on a strong correlation between population density and neighborhood vibrancy (Yue, 2016). A few reasons can explain this different result. The primary reason was that the study contains 23 block groups as a sample size, which is slightly small and may decrease statistical power. Following that, a few block groups with a high population density in small-sized residential areas have many lower numbers of Yelp reviews resulted from a lack of business activities, leading to a bias on certain block groups and even unreliable findings in the model. Therefore, it is not possible to input the socioeconomic variable as a control variable in the following models.

Table 11: Result of the Model 1

Model	R	R2	Adjusted R	Significant
1	0.336	0.113	0.071	0.117

Next, Table 12 shows the results of the multi-level modeling for the three major elements of urban form. In model 2, the two urban layout variables of urban form indicators, average block length and average block size were introduced into the regression model. The result demonstrates the urban layout variable only satisfy 20% of changes in urban vitality. Regardless of the low variability of the result, average block length has a negative correlation with urban vitality with a 90% level of confidence, while there was no significant evidence proving Jacobs' emphasis on a small block as an essential factor for vitality in the study (Jacobs, 1961).

When deciding to include street connectivity data of urban form indicators in model 3, the assumption is that the model will closely fit the data because of the increased goodness of fit (R2). Unfortunately, in comparing model 2 and model 3, R2 for the model 3 is 0.261, which is slightly higher than that for model 2. In other words, model 3 can only explain 26% of the variability in vitality. Meanwhile, except for the average block length, the significance of the rest independent variables is extremely higher than 0.05 and therefore the model 3 is not reliable, even though the significance of the model is 0.049 less than 0.05.

In addition to urban layout and street layout data, density information of urban form variables were included in model 4. The goodness of fit of the model dramatically increased to 0.598, more than two times of that of the model 2 and model 3. This implies that the density variable has a strong correlation with urban vitality. The results in the model show that both connected node ratio and activity density are positively associated with urban vitality. In conclusion, out of the three major elements of the urban fabric, density tends to have the most significant association with urban vitality. However, it is difficult to predict which one of the remaining two elements has a closer correlation with city vitality. Table 12 shows a side by side model result comparison.

Table 12: Summary Statistics of the Multilevel Regression Model

	Model 1			**Model 2**			**Model 3**			**Model 4**		
Variable	Beta	T statistics	Sifnificancy	Beta	T statistics	Sifnificancy	Beta	T statistics	Sifnificancy	Beta	T statistics	Sifnificancy
Population density	0.336	1.637	0.117									
Average Block Length				-0.493	-2.573	0.018	-0.489	-2.147	0.046	-0.197	-1.076	0.297
Average Block Size				-0.271	-1.417	0.172	-0.035	-0.149	0.883	0.182	1.001	0.331
Sidewalk Density Index							0.199	0.757	0.459	-0.055	-0.268	0.792
Connected Node Raito							0.316	1.483	0.155	0.387	2.445	0.026
Activity Desntiy										0.663	4.011	0.001
R2	0.071			0.208			0.261			0.598		
Significant	0.117			0.037			0.049			0.001		
Dependent Variable	Yelp Reviews											
N	23											

Correlations between Each Urban Form Indicator and Urban Vitality

To study each independent variable's influence on urban vitality, a stepwise regression function in SPSS was run to further examine each urban form indicator individually. An OLS multivariate regression model was performed on the variables that only allow the model to achieve a 90% level of confidence. The results of the stepwise regression demonstrate that, out of the five urban form variables, three are associated with urban vitality at a 90% level of confidence. Consequently, in the multivariate model, Yelp review represents the dependent variable, and the three extracted urban form indicators, including average block length, connected node ratio and activity density represent the independent variables of the model. As it is shown in Table 13 below, the significance of the model is 0.000, it is concluded that the relationship between variables is linear and the regression model is reliable. The goodness of fit R2 in the model is 0.610, indicating that the three variables satisfy 61% of variations in urban vitality.

In conclusion, out of the six selected urban form variables, three are significantly associated with urban vitality. Meanwhile, different independent variables play different roles in urban vitality. The correlation coefficients among the three variables in the selected block groups are shown in Table 13. The findings show that the independent variables of activity density and connected node ratio have a positive correlation with urban vitality, while as expected, average block length plays a negative role in creating urban vitality. In other words, people are more likely to engage in a connected area where there are more real intersections and a place with a high density of those attractive uses. For the average block length, the higher the average block length of a block group, the less the intersections in the block groups, leading to a less connected place and hurting urban vitality. In addition, in comparison among the independent variables, the activity density has the highest impact on urban vitality and the average block length has the lowest impact on urban vitality. The final result of the regression model can be summarized into the formula below:

$$Y_{\text{YELP REVIEW}} = 0.566 * X_{\text{ACTIVITY DENSITY}} + 0.304 * X_{\text{CONNECTED NODE RATIO}} - 0.28 * X_{\text{AVERAGE BLOCK LENGTH}}$$

Table 13: Summary Statistics of the OLS Multivariate Regression Model

Variable	Beta	T Statistics	Significance
Constant	-	0.000	1.000
Average Block Length	-0.280	-2.061	0.053
Connected Node Ratio	0.304	2.197	0.041
Activity Density	0.566	4.105	0.001
R2	0.610	-	-
Significant	0.000	-	-

Based on the model, sidewalk density index and average block size have a limited association with urban vitality based on the study on the 23 block groups. However, the result for the average block size in the study is different from the results found by previous researchers that show small block size has a strong positive association with vitality (Ye, 2017; Long, 2017; Sung, 2015). For instance, Ye summarized his findings on block size affecting urban vitality by studying two major districts in Shenzhen in China, including Luohu and Futian districts. The total area of the two districts is 60.6 square miles, while this study focuses on 23 block groups of downtown areas in the City of Austin, the City of Dallas and the City of San Antonio. As a result, the different results are contributed to different sizes of focus areas between this study and others. Compared to Ye's study, the study area in this research is much smaller. In addition, such different results may also result from different forms of the built environment. This study only emphasizes on the compact downtown areas in Texas, while Ye studied the whole districts that include both urban and suburban fabrics. For the sidewalk density index, although there is a common expectation that higher sidewalk density index, representing a walkable environment, encourages pedestrian activities, it is surprising that there is no statistically significant relationship between sidewalk density index and urban vitality in the study.

Chapter 6: Limitations & Future Research

Limitations of the Research

There are certain limitations in defining Yelp reviews as a metric of urban vitality. Though Yelp reviews, one of the most popular social media platforms, reflect social interaction among people, the reviews on businesses are too subjective. There are biases of the group, age and occupation in the reviews (Lu, 2019). Obviously, those people who have a memorably good or bad experience may write reviews, while many people are not inclined to use it for several reasons. For example, some children, elders or low-income populations may not have electrical equipment to write reviews over the businesses they interact with, leading to an incomplete dataset on reviews. In terms of the technique to obtain total reviews, Yelp reviews were manually conducted because of limited skills in coding.

Another limitation is that the research only studies 23 bock groups in the City of Austin, the City of Dallas and the City of San Antonio due to time and funding. Such a small-sized sample, which has limited statistical power, could result in different results from previous studies. The research only conducted the study in the three main cities, there is also a limitation to conclude how much we apply the results for other cities in the Texas Triangle. The study could be an exploratory analysis and starting point for future studies.

Future Research

Due to these limitations, instead of defining one check-in data from Yelp, future research should scrape social media data from various platforms, such as Twitter, and Instagram, to represent urban vitality and run multiple regression models to compare and verify each of them has a similar correlation with urban form variables. This complete set of social media data will provide a more robust analysis of how many activities took place through sharing on social media platforms and a more reliable result on the relationship between urban form and vitality. To conduct a more comprehensive study on urban vitality, in addition to social media data, GPS tracked mobile phone data, point of interest density, and pedestrian & vehicle counts should also be included in future research.

To collect the data, future research should rather automatically access and download location-based posts with coordinates from social media websites by web scraping and developing a Python code. This technique will not only provide more accurate datasets, but it also saves time on collecting data. Moreover, it is needed to increase the sample size in future research. Including surrounding areas of downtowns in the three main cities in the Texas Triangle will add more block groups and therefore effects of the study may be more statistically significant. If possible, the City of Houston should be included in further study as it plays an important role in the Texas Triangle.

Chapter 7: Conclusion

In conclusion, this study offers an exploratory analysis of how impacts of urban form of the built environment effect urban vitality in the Texas Triangle by using total Yelp reviews on a full category of businesses as a proxy for vitality in block groups. In addition to socioeconomic independent variables, six metrics were identified and, categorized into three elements of urban form that measure the variations in physical forms among 23 block groups in the Texas Triangle. A multi-level and multivariate regression model were applied to separately explore the degree each element of urban form correlates with urban vitality and each metric of urban form associates with urban vitality.

Overall, although based on the study on the small sample size, some of the metrics including average block size and sidewalk density index have a limited significant impact on urban vitality defined by the total number of Yelp reviews per block groups, which represents a conflicting result with previous studies, the implications of the regression models demonstrate that three metrics are significantly associated with urban vitality. Activity density and connected node ratio play a significantly positive role in urban vitality. Moreover, small block length is another essential factor to urban vitality. It is concluded with a negative correlation with urban vitality. Out of the three metrics, activity density has the most significant association with vitality.

Even though the study is exploratory, it provides some implications for urban planning and design practices. This study suggests that shortening block length is an effective measure to promote pedestrian activities and further improve urban vitality. As a result, future development in a small block length area will be advocated by the government, planners, designers and developers. Furthermore, in order to propose a vibrant place, real intersections should be encouraged along with small block length in the cities' future development plan, which will improve street connectivity and therefore traffic and pedestrian flow. Finally, with the positive effect of activity density on vitality, decision-makers should employ more human-oriented uses and diverse choices for different levels of the urban population.

References

Jacobs, J. 1961. “The Death and Life of Great American Cities.” *New York, NY: Random House.*

Montgomery, J. 1998. “Making a city: Urbanity, Vitality and Urban Design.” *Journal of Urban Design.*

Gehl, J. 1971. “Life Between Buildings: Using Public Space.” *New York: Van Nostrand Reinhold.*

Lynch, K. 1981. “Theory of Good City Form.” *Massachusetts: MIT Press.*

Katz, P. 1993. “The New Urbanism: Toward an Architecture of Community.” *McGraw-Hill Professional: New York, NY.*

Ye, Y. & Li, D. & Liu, X. 2017. “How Block Density and Typology Affect Urban Vitality: An Exploratory Analysis in Shenzhen, China.” *Journal of Urban Geography, Vol 39, 2018 Issue 4.*

Lu, S. & Huang, Y. & Shi, C. & Yang, X. 2019. “Exploring the Associations Between Urban Form and Neighborhood Vibrancy: A Case Study of Chengdu China.” *ISPRS Int. J. Geo-Inf. 2019, 8(4), 165*; https://doi.org/10.3390/ijgi8040165

He, Q. 2018. “The Impact of Urban Growth Patterns on Urban Vitality in Newly Built-Up Areas Based on An Association Rules Analysis Using Geographical ‘Big Data’.” *Land Use Policy Vol 78, November 2018, Pages 726-738*

Yue, Y. 2016. “Measurements of POI-based Mixed Use and Their Relationships with Neighborhood Vibrancy.” *International Journal of Geographical Information Science, Vol 31, 2017-Issue 4*

Sulis, P. & Manley, E. & Zhong, C. & Batty, M. 2018. “Using Mobility Data as Proxy for Measuring Urban Vitality.” *Journal of Spatial Information Science, Number 16, Page 137-162*

Zumelzu, A. & Barrientos-Trinanes, M. 2019. “Analysis of the Effects of Urban Form on Neighborhood Vitality: Five Cases in Valdivia, Southern Chile.” *Journal of Housing and the Built Environment, Pages 1-29*

Wu, J. & Ta, N. & Song, Y. & Lin, J. & Chi, Y. 2018. “Urban Form Breeds

Neighborhood Vibrancy: A Case Study Using A GPS-based Activity Survey in Suburban Beijing." *Cities, Vol 74, Pages 100-108.*

Long, Y & Huang, C. 2017. "Does Block Size Matter? The Impact of Urban Design on Economic Vitality for Chinese Cities." https://doi-org.ezproxy.lib.utexas.edu/10.1177/2399808317715640

Zarin, S. & Niroomand, M. & Heidari, A. 2015. "Physical and Social Aspects of Vitality Case Study: Traditional Street and Modern Street in Tehran." *Procedia-Social and Behavioral Sciences, Vol. 170, Pages 659-668*

Kitchin, R. 2014. "The Real-time City? Big Data and Smart Urbanism." *Geo Journal, Vol. 79, Issue 1, Pages 1-14.*

Drewes, J.E. & Aswegen, M. 2010. "Determining the Vitality of Urban Centers." *WIT Transportation on Ecology and the Environment, Vol. 142, WIT Press.*

Anderson, W. & Kanaroglou, P. & Miller, E. 1996. "Urban Form, Energy and the Environment: A Review of Issues, Evidence and Policy." *Urban Studies, Vol. 33, No. 1, Pages 7-35*

Texas Triangle. American 2050.

Dempsey, N. & Brown, C. 2009. "Elements of Urban Form." *DOI: 10.1007/978-1-4020-8647-2_2*

Koohsari, M. & Sugiyama, T. & Lamb, K.E. & Villanueva, K. & Owen, N. 2014. "Street Connectivity and Walking for Transport: Role of Neighborhood Destinations." *Preventive Medicine, Vol. 66, Pages 118-122*

Venerandi, A. 2017. "A Quantitative Method to Study the Relationship between Urban Form and City Livability Indexes." *University College London.*

Sung, H. & Lee, S. 2015. "Residential Built Environment and Walking Activity: Empirical Evidence of Jane Jacobs' Urban Vitality." *Transportation Research Part D: Transport and Environment, Vol. 41, Pages 318-329*